

Importance of Selecting Proper Basis Set in Quantum Mechanical Studies of Potential Energy Surfaces of Carbohydrates

JENN-HUEI LII,¹ BUYONG MA,² NORMAN L. ALLINGER¹

¹Computational Center for Molecular Structure and Design, Department of Chemistry, University of Georgia, Athens, Georgia 30602-2556

²Laboratory of Experimental and Computational Biology, NCI-FCRDC, Frederick, Maryland 21702

Received 22 June 1999; accepted 22 June 1999

ABSTRACT: An extensive quantum mechanical study of a water dimer suggests that the introduction of a diffuse function into the basis set, which significantly reduces the basis set superposition error (BSSE) in the hydrogen bonding energy calculation, is the key to better calculations of the potential energy surfaces of carbohydrates. This article examines the potential energy surfaces of selected D-aldo- and D-ketohexoses (a total of 82 conformers) by quantum mechanics (QM) and molecular mechanics (MM) methods. In contrast to the results with a smaller basis set (B3LYP/6-31G** 5d), we found at the higher level calculation (B3LYP/6-311++G(2d,2p)/B3LYP/6-31G** 5d) that, in most cases, the furanose forms are less stable than the pyranose forms. These discrepancies are mainly due to the fact that intramolecular hydrogen bonding energies are overestimated in the lower level calculations. The higher level QM calculations of the potential energy surfaces of D-aldo- and D-ketohexoses now are more comparable to the MM3 results. © 1999 John Wiley & Sons, Inc. *J Comput Chem* 20: 1593–1603, 1999

Keywords: *ab initio*; diffuse function; basis set superposition error; molecular mechanics; water dimer; hydrogen bonding; carbohydrates; aldohexose; ketohexose

Correspondence to: N. L. Allinger; e-mail: allinger@sunchem.chem.uga.edu

Contract/grant sponsor: Tripos Associates

Introduction

It has long been a challenging task for theoretical chemists to get a clear picture of the potential surface of carbohydrates. Even a monosaccharide like glucose in which the pyranose form dominates the conformational distribution in aqueous solution possess a tremendous challenge to theory, because of the complexity of its conformational energy surface. Although the conformational analysis of monosaccharides has been extensively explored by a variety of computational methods,¹⁻⁶ the factors that determine the preferred structures of monosaccharides are still not very well understood. The complexity of these conformational energy surfaces is believed to be largely due to a combination of the anomeric effects, the dipole-dipole interactions, and the high degrees of hydrogen bonding (internal and external) present in monosaccharides.

The anomeric effect, which stabilizes the axial anomer of a pyranose, is known to result mainly from the hyperconjugative delocalization of the ring oxygen lone pair electrons into the anti-exocyclic $\sigma_{C(1)-O}^*$ orbital. Its role in the conformational analysis of monosaccharides has been well recognized and studied.⁷⁻¹³ Although the anomeric effect is very important in determining the conformational equilibria between α and β anomers, its energy contribution to the conformational energy surface of monosaccharides, compared to that of intramolecular hydrogen bonding, is relatively small. In glucose, for instance, there are four possible intramolecular hydrogen bonding interactions, but only one anomeric effect. If we take 3 kcal/mol (from water dimer) for each hydrogen bonding interaction and 1 kcal/mol (ref. 14) for the anomeric effect, then their energy contribution ratio in the gas phase is about 12:1. Therefore, it is the intramolecular hydrogen bonding and not the anomeric effect that mainly determines the conformational energy surfaces of monosaccharides, and it needs to be examined more extensively. In this article we use density functional theory (DFT) to examine the hydrogen bonding interactions of the water dimer more carefully. Earlier work¹⁵ showed unexpectedly rather large discrepancies between *ab initio* and DFT calculations on the one hand and molecular mechanics (MM3) calculations on the other. The reasons for the discrepancies were unclear and led to our examining both the impor-

tance of polarization in MM and the use of a better basis set (including a diffuse function) in DFT. The latter topic proved to be more important and will be discussed here. The major discrepancy between DFT and MM is removed when the DFT calculations are improved with the larger basis set. The polarization problem will be discussed in a separate article.

Computational Methods

All quantum mechanics (QM) calculations were carried out with the Gaussian 94 program.¹⁶ The linear water dimer was fully optimized using second-order Møller-Plesset perturbation theory (MP2) and DFT Becke three-parameter Lee-Yang-Parr (B3LYP), using various basis sets up to AUG-cc-pVQZ (344 basis functions). A total of 64 D-aldohexose and 18 D-ketohexose conformers were optimized at the B3LYP/6-31G** level with five d-type functions added on the carbon and oxygen atoms.¹⁷ The optimized structures were then used for the B3LYP/6-311++G(2d,2p) single-point calculations. In the water dimer cases the basis set superposition errors (BSSEs) that arise from the compensation of inadequacies of the basis set by diffuse functions of neighboring centers were corrected by the counterpoise method of Boys and Bernardi.¹⁸

The MM calculations were carried out with the MM3(00)¹⁹ program using the default dielectric constant value of 1.5, which is standard for the gas phase.

Results and Discussion

As we mentioned previously, the energy contribution of intramolecular hydrogen bonding to the conformational energy surface of the monosaccharides was found to be so significant that we could not ignore the errors, especially the BSSEs, that might be introduced in the low level QM calculations. Unless we can find a good way to correct the intramolecular BSSE, we have to search for a better and practical basis set, which has a smaller BSSE effect, for the calculations of monosaccharides. Although Brickmann et al.²⁰ proposed a procedure to correct the intramolecular BSSE in their 1,2-ethanediol study, the procedure seems to be too complicated for our purpose. Therefore, in the

present work we focused instead on finding a better basis set for our carbohydrate studies.

In the next two sections we discuss some detailed investigations of hydrogen bonding interactions between hydroxyl groups, and then we extend the discussion to the conformational energy surfaces of hexoses.

WATER DIMER AND HYDROGEN BONDING ENERGY

To see how seriously the BSSE might affect the calculations of the conformational energy surfaces of monosaccharides, we started with a deeper exploration of hydrogen bonding interactions between hydroxyl groups using the water dimer as a model. Because the geometric and energetic properties of the water dimer have been well studied experimentally and theoretically^{21,22} and all of the research agrees that the linear conformer is the most stable one, we chose the linear water dimer as the focus for our study.

We may systematically compare the results of the linear water dimer calculated at different levels of theory up to B3LYP/AUG-cc-pVQZ (344 basis functions). As one can see from Table I, the binding energies of the linear water dimer calculated at different levels of theory were quite scattered (from -4.6 to -7.5 kcal/mol) before the BSSE corrections were applied. However, the energies gradually converged, as one would expect, as the basis set size increased. The data in Table I and Figure 1 clearly show that the binding energies at all levels of calculation approached the value of -4.58 kcal/mol, which was calculated with the highest basis set AUG-cc-pVQZ, after the BSSE corrections were applied. These results also confirm our belief that the BSSEs correction is crucial for hydrogen bonding calculations when an inadequate basis set is used.

Another important thing we would like to point out is that the BSSEs were significantly reduced when the diffuse functions were introduced into the basis sets (see Fig. 1). Without diffuse functions included, even in the calculation with a big basis set like cc-pVQZ (230 basis functions), the BSSEs were still significant (-0.77 kcal/mol). We also found that the water dimer geometries were always poorly calculated when no diffuse functions were included (see Table I). Therefore, the inclusion of diffuse functions in the basis set, as well as the BSSE correction, are likely the keys to better calculations of hydrogen bonding interactions in a water dimer and, by implication, of

hydrogen bonding in general and carbohydrates in particular.

Based on the results shown in the Table I and Figure 1, we can clearly see that the best level of theory for hydrogen bonding calculations (in terms of accuracy) that we examined is B3LYP/AUG-cc-pVQZ for the water dimer. Unfortunately, the basis set is too big for practical use. On the other hand, two popular levels of theory, B3LYP/6-31G** and MP2/6-31G**, give rather poor geometric and energetic results in the water dimer calculations. Their BSSEs are more than 30% of their uncorrected binding energies. Therefore, the practical choice seems to be the B3LYP/6-311++G(2d,2p) theory, because its size is moderate (94 basis functions) and it gives a relative small BSSE (-0.35 kcal/mol, 7% of the uncorrected binding energy), as well as a reasonably good water dimer geometry.

Our MM3 calculation for the binding energy of the linear water dimer is in fair agreement with the experiment and QM results (Table I).²³ However, its geometry is rather poorly calculated. The angle of $\text{O}\cdots\text{O}\cdots\text{X}$ ($\text{O}\cdots\text{X}$ is the bisector of $\text{H}-\text{O}-\text{H}$) in the linear water dimer was calculated to be 160° , which is too big compared to the experimental value of 123° (10) and the best QM value of 124° (see Table I). We believe this discrepancy is mainly due to the lack of lone-pair directionality in the MM3 hydrogen bonding formulation,²⁴ and recent studies²⁵ confirm this. We admit this shortcoming in the MM3 force field might be serious in the conformational energy surface calculation of systems like carbohydrates in which high levels of hydrogen bond functionality are present, and we will deal with this problem separately.

CONFORMATIONAL ENERGY SURFACES OF HEXOSES

In our previous study¹⁵ we systematically compared the DFT (at B3LYP/6-31G** 5d theory) and the MM3(96)²⁶ results of the conformational energies of eight D-aldohexoses and four D-ketohexoses. However, the comparison was disappointing. The average discrepancies for the total energies between the two methods were on the order of 1.5 kcal/mol, and the relative energies for anomeric pairs calculated with MM3(96) and B3LYP had uniformly opposite signs. The MM3 results appeared to be so poor that they raised questions about the transferability of our alcohol/ether force field (which was developed years ago²⁷) to carbohydrates. We also initially speculated on the accu-

TABLE I.
Binding Energies and Geometries for Linear Water Dimer at Selected Levels of Theory.

(H ₂ O) ₂ Linear, C _s	Exp.	MM3(00)	RHF / 6-31G**	B3LYP / 6-31G**	MP2 / 6-31G**	B3LYP / 6-31++G**
Basis functions			50	50	50	62
A + B			−152.05605958	−152.85150064	−152.45080241	−152.87783418
A ₀ + B ₀			−152.04722998	−152.83947334	−152.43957152	−152.86824725
A + B(0)			−76.02376836	−76.42009146	−76.22020865	−76.43435886
B + A(0)			−76.02499730	−76.42328018	−76.22315249	−76.43516017
E (uncorrected)			−5.541	−7.547	−7.047	−6.016
BSSE			−0.964	−2.446	−2.378	−0.798
E (corrected)	−5.21(1.5) ^b	−4.770	−4.577	−5.101	−4.669	−5.218
Dipole moment	2.6 ^c	3.508	2.667	1.807	1.956	2.842
Imaginary frequency		0	0	1 ^f	0	N/A ^g
O ... O distance	2.976 ^d	2.942	2.981	2.875	2.912	2.883
O ... H distance		2.007	2.038	1.931	1.966	1.920
< O ... O — H	6 (20) ^d	8.9	5.0	11.5	9.8	6.6
< O ... O ... X ^e	123 (10) ^d	160.3	117.9	92.8	99.3	123.6

(H ₂ O) ₂ Linear, C _s	MP2 / 6-31++G**	BLYP / 6-311G(2d,2p)	B3LYP / 6-311++G(2d,2p) ^a	MP2 / 6-311++G(2d,2p)	B3LYP / cc-pVQZ	B3LYP / AUG-cc-pVQZ
Basis functions	62	82	94	94	230	344
A + B	−152.47696437	−152.91586251	−152.93199325	−152.60113514	−152.94777339	−152.95177412
A ₀ + B ₀	−152.46675187	−152.90423495	−152.92408000	−152.59258943	−152.93929207	−152.94448081
A + B(0)	−76.23398471	−76.45271537	−76.46215380	−76.29666069	−76.46989813	−76.47219052
B + A(0)	−76.23534182	−76.45644303	−76.46248683	−76.29732748	−76.47061998	−76.47227070
E (uncorrected)	−6.048	−7.296	−4.966	−5.363	−5.322	−4.577
BSSE	−1.616	−3.090	−0.352	−0.878	−0.769	0.012
E (corrected)	−4.793	−4.207	−4.614	−4.485	−4.553	−4.589
Dipole moment	3.281	1.969	2.633	2.783	2.559	2.633
Imaginary frequency	N/A	N/A	0	0	N/A	N/A
O ... O distance	2.911	2.905	2.918	2.917	2.913	2.918
O ... H distance	1.944	1.952	1.955	1.956	1.950	1.959
< O ... O — H	3.4	8.0	5.2	4.6	4.7	6.2
< O ... O ... X ^e	133.0	100.2	120.2	123.7	117.3	124.0

All QM energies are in Hartrees; MM3 steric energies are in kilocalories per mole; *E* (uncorrected), BSSE, and *E* (corrected) are in kilocalories per mole; dipole moments are in debyes; all distances are in Angstroms; all angles are in degrees.
^aAn approximation to this method [B3LYP / 6-311++G(2d,2p) // B3LYP / 6-31G** 5d] is used in the present sugar work.
^bReference 20(a).
^cReferences 20(b) and 20(c).
^dReference 20(d).
^eO ... X is the bisector of H—O—H.
^fOne negative frequency is about −18 cm^{−1}.
^gN/A, no frequency calculation was carried out.
^hA + B: Energy of dimer A − B; A₀ + B₀: Sum of energies of isolated monomers A and B; A + B(0): Energy of monomer A with empty orbital of B; B + A(0): Energy of monomer B with empty orbital of A.

racy of the B3LYP/6-31G** 5d results.¹⁵ After the studies of the hydrogen bonding interactions between hydroxyl groups described in the previous section, we had some evidence to support some of our previous speculations.
According to our present study, the B3LYP/6-31G** 5d theory used in our previous study gave poor results for the water dimer calculations (see Table I). It not only gave a geometry with poor angles (which however are constrained in carbohydrates, see Table I), but also a large BSSE (−2.4

kcal/mol) and an imaginary frequency (−18 cm^{−1}). The BSSE is so big (32% of the uncorrected binding energy) that we cannot ignore the possibility of potentially serious errors in the intramolecular hydrogen bonding energies in the calculations on monosaccharides. In the present work we went over the calculations of the D-hexoses again with a higher level of theory, the B3LYP/6-311++G(2d,2p), which was shown to give a better geometry and a smaller BSSE in our water dimer study (see Table I). Because there is a total of 444 basis

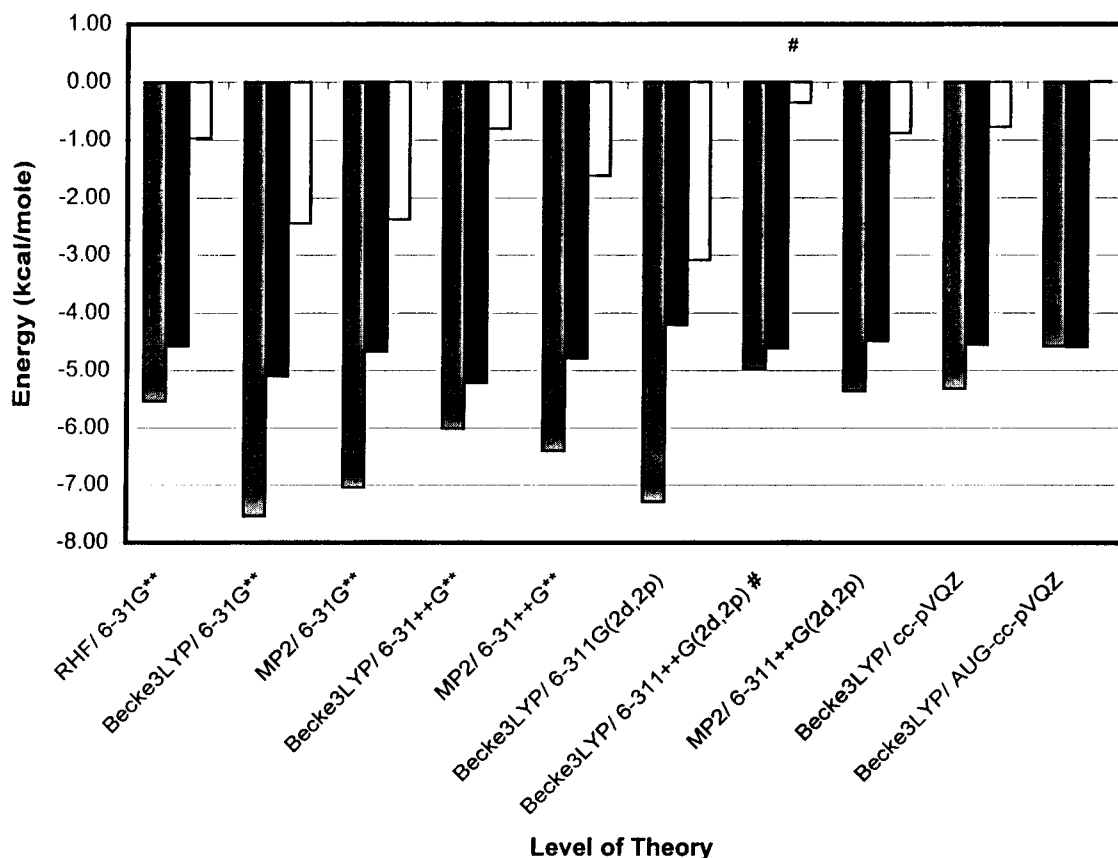


FIGURE 1. Binding energies and BSSE of linear water dimer at selected levels of theory: (■) binding energy (uncorrected), binding energy (BSSE corrected), and (□) BSSE. (#) An approximation to this method [B3LYP/6-311++G(2d,2p)//B3LYP/6-31G** 5d] is used in the present sugar work.

functions for each calculation here, it would be a very CPU-time intensive process to fully optimize all 82 conformers. Hence, it appeared that an alternative was desirable. Based on our preliminary results (see Table II) we found that the calculated relative energies among the selected D-hexose conformers were very similar by either of the two methods: a full optimization at the B3LYP/6-311++G(2d,2p) level or a single-point calculation using this level of theory with a B3LYP/6-31G** 5d geometry. These results show that the geometries of D-hexoses are not much different by either method. Therefore, the single-point energy calculation at the B3LYP/6-311++G(2d,2p) level using the B3LYP/6-31G** 5d geometry seemed to be an expedient choice.

As we suspected, the conformational energy surfaces of D-hexoses calculated at the higher level of theory were significantly different than those found in our previous study.¹⁵ In contrast to our previous finding we found that the furanose forms are mostly less stable than the pyranose forms.

From Table III we can see that the total relative energies of the furanoses sharply increased (2–4 kcal/mol), and the energy gap between the α and β anomers significantly decreased (by 0.8 kcal/mol on the average, see Table IV). As we pointed out earlier,¹⁵ the furanoses were stabilized relative to the pyranoses by their strong hydrogen bonding. But it is now clear that the hydrogen bonding energies were overestimated by the small basis set, and the intramolecular hydrogen bonding energies were significantly reduced in the B3LYP/6-311++G(2d,2p)//B3LYP/6-31G** 5d calculations. The net result was a general shift of stability from the furanoses to the pyranoses.

Comparing the older and newer sets of calculations, the root mean square (RMS) and signed average errors of the total energy difference between the MM3(00) and DFT results are reduced from 3.29 to 1.76 and 1.35 to 0.20 kcal/mol, respectively. In other words, the discrepancies between the earlier DFT and MM3 results were largely due to the inaccuracy in the DFT calculations that re-

TABLE II.
Relative Energies of Glucose, Allose, and Altrose Conformations at Different Levels of Theory.

Conformer		B3LYP / 6-31G** 5d		B3LYP / 6-311++G(2d,2p) // B3LYP / 6-31G** 5d		B3LYP / 6-311++G(2d,2p)	
		α	β	α	β	α	β
		ΔE	ΔE	ΔE	ΔE	ΔE	ΔE
Glucose							
Pyranose	⁴ C ₁ GG	0.55	2.24	-0.02	0.76	-0.08	0.74
	⁴ C ₁ GT	0.81	2.77	-0.12	1.26	-0.16	1.28
	⁴ C ₁ TG	0.00	1.82	0.00	0.91	0.00	0.95
Furanose	GGGG	2.57	-0.23	4.69	3.61	4.50	3.34
Allose							
Pyranose	⁴ C ₁ GG	-0.93	0.50	0.59	0.44	0.44	0.35
	⁴ C ₁ GT	-0.63	0.94	0.50	0.46	0.39	0.40
	⁴ C ₁ TG	-1.58	0.00	0.66	0.65	0.56	0.59
Furanose	GGGG	-3.33	-1.03	-0.40	0.87	-0.63	0.43
Altrose							
Pyranose	⁴ C ₁ GG	0.71	2.89	2.20	2.48	2.06	2.38
	⁴ C ₁ GT	0.80	3.64	2.02	3.24	1.92	3.05
	⁴ C ₁ TG	-0.29	2.56	2.12	3.11	2.00	3.06
Furanose	GGGG	3.10	-1.42	3.80	1.08	3.51	0.72

All relative energies, ΔE , are in kilocalories per mole. The conformation assigned the zero of energy is arbitrary and was chosen as the lowest energy form of glucose (pyranose) with the small basis set calculation.

sulted from the lack of diffuse functions in the basis set. The RMS and signed average errors in the anomeric energy difference, $E_{\beta-\alpha}$, were also significantly improved from 3.33 to 2.15 and -1.99 to -1.23 kcal/mol, respectively. Compared to the new DFT results, however, MM3(00) still predicts β anomers to be more stable in most cases (see Table IV, Fig. 2). The sources of these discrepancies are still unclear to us at this moment, and we are continuing a further investigation into this problem. Nevertheless, our MM3(00) calculations are now much more comparable to the higher level DFT results and the improvements here are strictly due to the inadequate basis set used in our previous study.

Conclusions

Intramolecular hydrogen bonding was found to be the key factor that determines the conformational energy surfaces of monosaccharides. The present study shows that the neglect of the intramolecular BSSE corrections is the likely cause of the poor results for energy surfaces of monosaccharides from low level QM calculations. Unless intramolecular BSSE can be properly taken care of,

a better theory, which gives smaller intramolecular BSSEs, is required in the calculations of energy surfaces of systems like carbohydrates. Our extensive QM study of a water dimer suggests that the introduction of a diffuse function into the basis set, which significantly reduces the BSSE in the hydrogen bonding energy calculation, is the key to the better calculation of the potential surface of carbohydrates.

This work demonstrates a dramatic change in the QM results of the potential surfaces of selected D-aldo- and D-ketohexoses (a total of 82 conformers) when a better theory is used, such as the B3LYP/6-311++G(2d,2p) single-point calculation using B3LYP/6-31G** 5d geometry. The QM results at the higher level of theory are now much more comparable to our MM3(00) results than those of the previous calculation¹⁵ at the B3LYP/6-31G** 5d level of theory. Unlike our previous finding, the furanose forms are generally less stable than the pyranose forms in the gas phase, which is known to be the general preference of pyranose rings in solution. This work shows the importance of high accuracy in the QM calculations. Although the hydrogen bonding energy is calculated to be too strong by only 1 kcal/mol with the small basis set, because several hydrogen

TABLE III.
Total and Relative Energies of Selected Monosaccharides.

Conformer	B3LYP / 6-31G** 5d			B3LYP / 6-311++G(2d,2p) // B3LYP / 6-31G** 5d			MM3(00)		
	α		β	α		β	α		β
	E	ΔE	E	E	ΔE	E	E	ΔE	E
Glucose									
Pyranose									
⁴ C ₁ GG	-687.17792039	0.55	-687.17522264	-687.42637965	-0.02	-687.42514319	-265.44	0.65	-266.05
⁴ C ₁ GT	-687.17750351	0.81	-687.17438684	-687.42655326	-0.12	-687.42434463	-266.16	-0.07	-265.15
⁴ C ₁ TG	-687.17879624	0.00	-687.17590043	-687.42635484	0.00	-687.42490470	-266.09	0.00	-266.58
Furanose									
GGGG	-687.17469979	2.57	-687.17916281	-687.41887943	4.69	-687.42059897	-263.22	2.87	-264.53
Allose									
Pyranose									
⁴ C ₁ GG	-687.18028494	-0.93	-687.17799425	-687.42541794	0.59	-687.42565957	-262.82	3.27	-264.90
⁴ C ₁ GT	-687.17980250	-0.63	-687.17730511	-687.42556500	0.50	-687.42562659	-263.84	2.25	-265.75
⁴ C ₁ TG	-687.18131179	-1.58	-687.17879376	-687.42530979	0.66	-687.42531411	-263.71	2.38	-265.61
Furanose									
GGGG	-687.18409870	-3.33	-687.18043962	-687.42699777	-0.40	-687.42497141	-264.80	1.29	-266.56
Altrose									
Pyranose									
⁴ C ₁ GG	-687.17765968	0.71	-687.17418679	-687.42284835	2.20	-687.42240477	-260.85	5.24	-264.46
⁴ C ₁ GT	-687.17752675	0.80	-687.17299556	-687.42313923	2.02	-687.42119440	-261.79	4.30	-264.08
⁴ C ₁ TG	-687.17925804	-0.29	-687.17471138	-687.42297976	2.12	-687.42139093	-261.82	4.27	-263.06
Furanose									
GGGG	-687.17385927	3.10	-687.18106476	-687.42029742	3.80	-687.42464043	-261.83	4.26	-264.59
Mannose									
Pyranose									
⁴ C ₁ GG	-687.17690132	1.19	-687.17592366	-687.42431919	1.28	-687.42370804	-263.11	2.98	-265.73
⁴ C ₁ GT	-687.17665064	1.35	-687.17563326	-687.42420238	1.35	-687.42411111	-263.61	2.48	-266.33
⁴ C ₁ TG	-687.17791814	0.55	-687.17741749	-687.42409033	1.42	-687.42413396	-263.62	2.47	-266.24
Furanose									
GGGG	-687.17511061	2.31	-687.18332161	-687.41957104	4.26	-687.42201626	-262.21	3.88	-260.52
Gulose									
Pyranose									
⁴ C ₁ GG	-687.17981978	-0.64	-687.17886325	-687.42399349	1.48	-687.42498592	-263.64	2.45	-265.61
⁴ C ₁ GT	-687.17238943	4.02	-687.17152672	-687.41934544	4.40	-687.42063540	-262.32	3.77	-263.65
⁴ C ₁ TG	-687.17285952	3.73	-687.17079071	-687.41881826	4.73	-687.41877902	-259.72	6.37	-260.89
Furanose									
GGGG	-687.18224814	-2.17	-687.17328269	-687.42163378	2.96	-687.41694018	-263.83	2.26	-260.68
Idose									
Pyranose									
⁴ C ₁ GG	-687.18346470	-2.93	-687.17641757	-687.42599834	0.22	-687.42088150	-264.69	1.40	-264.12
⁴ C ₁ GT	-687.17604733	1.72	-687.17425184	-687.42119306	3.24	-687.42094576	-263.34	2.75	-264.76
⁴ C ₁ TG	-687.17449151	2.70	-687.17435823	-687.41891749	4.67	-687.41978504	-260.62	5.47	-262.73
Furanose									
GGGG	-687.18021390	-0.89	-687.17468429	-687.419300003	4.43	-687.41853054	-262.58	3.51	-263.22

(Continued)

TABLE III.
(Continued)

	B3LYP/6-31G** 5d				B3LYP/6-311++G(2d,2p)//B3LYP/6-31G** 5d				MM3(00)			
	α		β		α		β		α		β	
	E	ΔE	E	ΔE	E	ΔE	E	ΔE	E	ΔE	E	ΔE
Galactose												
Pyranose												
4C ₁ GG	-687.17842454	0.23	-687.17501386	2.37	-687.42480467	0.97	-687.42221400	2.60	-266.04	0.05	-264.94	1.15
4G ₁ GT	-687.17835211	0.28	-687.17505425	2.35	-687.42594635	0.26	-687.42353613	1.77	-266.17	-0.08	-265.36	0.73
4C ₁ TG	-687.17818580	0.38	-687.17491384	2.44	-687.42456470	1.12	-687.42227044	2.56	-263.89	2.20	-262.92	3.17
Furanose												
GGGG	-687.18106477	-1.42	-687.17566431	1.97	-687.42464023	1.08	-687.42299332	2.11	-264.59	1.50	-263.17	2.92
Talose												
Pyranose												
4C ₁ GG	-687.1800010	-0.76	-687.17322703	3.49	-687.42378005	1.62	-687.41819527	5.12	-265.74	0.35	-266.39	-0.30
4C ₁ GT	-687.1805360	-1.09	-687.17819376	0.38	-687.42526939	0.68	-687.42367553	1.68	-266.15	-0.06	-266.91	-0.82
4C ₁ TG	-687.1800447	-0.78	-687.17847562	0.20	-687.42383207	1.58	-687.42279202	2.24	-263.69	2.40	-264.97	1.12
Furanose												
GGGG	-687.1804401	-1.03	-687.18409874	-3.33	-687.42494396	0.89	-687.42671353	-0.23	-266.56	-0.47	-264.80	1.29
Fructose												
Pyranose												
2C ₅ GG	-687.17051459	5.20			-687.41633670	6.29			-259.85	6.24		
2C ₅ GT	-687.17328693	3.46			-687.42013840	3.90			-262.91	3.18		
2C ₅ TG												
5C ₂ GG	-687.17974926	-0.60	-687.16741231	7.14			-687.41137749	9.40			-255.43	10.66
5C ₂ GT	-687.18258096	-2.37	-687.16953390	5.81			-687.41621777	6.36			-258.69	7.40
Furanose												
GGGG	-687.18463013	-3.66	-687.17649377	1.44	-687.42345681	1.82	-687.42250200	2.42	-262.47	3.62	-260.72	5.37
Tagatose												
Pyranose												
5C ₂ GT	-687.18274218	-2.48	-687.17194587	4.30	-687.42740679	-0.66	-687.41808280	5.19	-265.24	0.85	-261.94	4.15
Furanose												
GGGG	-687.18302891	-2.66	-687.19008224	-7.08	-687.42245022	2.45	-687.42944079	-1.94	-260.47	5.62	-261.88	4.21
Psicose												
Pyranose												
5C ₂ GT	-687.18769484	-5.58	-687.17328651	3.46	-687.43058364	-2.65	-687.41965568	4.20	-266.21	-0.12	-262.31	3.78
Furanose												
GGGG	-687.18207634	-2.06	-687.18160169	-1.76	-687.42503203	0.83	-687.42364990	1.70	-263.12	2.97	-260.92	5.17
Sorbose												
5G ₂ GT	-687.18514572	-3.98	-687.17139053	4.65	-687.43118281	-3.03	-687.42018143	3.87	-268.39	-2.30	-264.26	1.83
Furanose												
GGGG	-687.17827367	0.33	-687.18264465	-2.41	-687.42324129	1.95	-687.42191680	2.78	-260.20	5.89	-261.64	4.45
RMS Errors of												
$\Delta E_{\text{MM3}} - \Delta E_{\text{B3LYP/6-31G** 5d}}$											3.29 (avg. 1.35)	
$\Delta E(\beta - \alpha)_{\text{MM3}} - \Delta E(\beta - \alpha)_{\text{B3LYP/6-31G** 5d}}$											3.33 (avg. -1.99)	
$\Delta E_{\text{MM3}} - \Delta E_{\text{B3LYP/6-311++G(2d,2p)//B3LYP/6-31G** 5d}}$											1.76 (avg. 0.20)	
$\Delta E(\Delta - \alpha)_{\text{MM3}} - \Delta E(\beta - \alpha)_{\text{B3LYP/6-311++G(2d,2p)//B3LYP/6-31G** 5d}}$											2.15 (avg. -1.23)	

MM3 energies are actually heats of formation and are in kilocalories per mole; all QM energies are equilibrium energies and are in Hartrees; all relative energies, ΔE , are in kilocalories per mole.

TABLE IV.
Conformational Energy Difference between β - and α -Anomers ($E_{\beta} - E_{\alpha}$) of Selected Hexoses.

		B3LYP / 6-311++G(2d,2p) // B3LYP / 6-31G** 5d			MM3(00)
No.	Conformer	B3LYP / 6-31G** 5d			
1	Glucose	P- ⁴ C ₁ GG	1.69	0.78	−0.61
2		P- ⁴ C ₁ GT	1.96	1.39	1.01
3		P- ⁴ C ₁ TG	1.82	0.91	−0.49
4	Allose	F-GGGG	−2.80	−1.08	−1.31
5		P- ⁴ C ₁ GG	1.44	−0.15	−2.08
6		P- ⁴ C ₁ GT	1.57	−0.04	−1.91
7		P- ⁴ C ₁ TG	1.58	0.00	−1.90
8	Altrose	F-GGGG	2.30	1.27	−1.76
9		P- ⁴ C ₁ GG	2.18	0.28	−3.61
10		P- ⁴ C ₁ GT	2.84	1.22	−2.29
11		P- ⁴ C ₁ TG	2.85	1.00	−1.24
12	Mannose	F-GGGG	−4.52	−2.73	−2.76
13		P- ⁴ C ₁ GG	0.61	0.38	−2.62
14		P- ⁴ C ₁ GT	0.64	0.06	−2.72
15		P- ⁴ C ₁ TG	0.31	−0.03	−2.62
16	Gulose	F-GGGG	−5.15	−1.53	1.69
17		P- ⁴ C ₁ GG	0.60	−0.62	−1.97
18		P- ⁴ C ₁ GT	0.54	−0.81	−1.33
19		P- ⁴ C ₁ TG	1.30	0.02	−1.17
20	Idose	F-GGGG	5.63	2.95	3.15
21		P- ⁴ C ₁ GG	4.42	3.21	0.57
22		P- ⁴ C ₁ GT	1.13	0.16	−1.42
23		P- ⁴ C ₁ TG	0.08	−0.54	−2.11
24	Galactose	F-GGGG	3.47	0.48	−0.64
25		P- ⁴ C ₁ GG	2.14	1.63	1.10
26		P- ⁴ C ₁ GT	2.07	1.51	0.81
27		P- ⁴ C ₁ TG	2.05	1.44	0.97
28	Talose	F-GGGG	3.39	1.03	1.42
29		P- ⁴ C ₁ GG	4.25	3.50	−0.65
30		P- ⁴ C ₁ GT	1.47	1.00	−0.76
31		P- ⁴ C ₁ TG	0.98	0.65	−1.28
32	Fructose	F-GGGG	−2.30	−1.11	1.76
33		F-GGGG	5.11	0.60	1.75
34	Tagatose	P- ⁵ C ₂ GT	6.77	5.85	3.30
35		F-GGGG	−4.43	−4.39	−1.41
36	Psicose	P- ⁵ C ₂ GT	9.04	6.86	3.90
37		F-GGGG	0.30	0.87	2.20
38	Sorbose	P- ⁵ C ₂ GT	8.63	6.90	4.13
39		F-GGGG	−2.74	0.83	−1.44
$E_{\beta} - E_{\alpha}$ Mean ^a			1.62	0.87	−0.37

All energies are in kilocalories per mole.

^aThis energy difference is the size of the anomeric effect. The experimental value is about 1 kcal/mol in solution. The small basis set gives too large a value, and the large basis set is about right. The MM3 value is small and of incorrect sign.

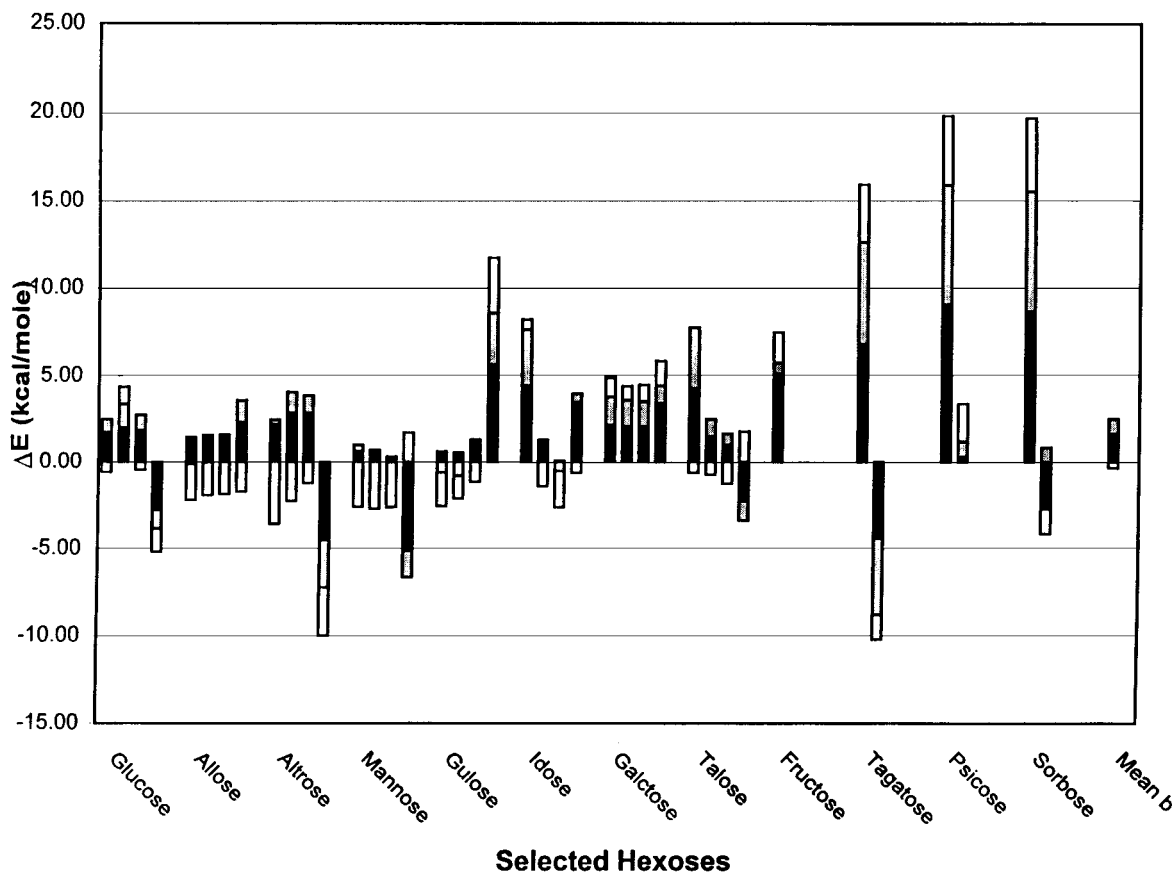


FIGURE 2. Conformational energy difference between β - and α -anomers ($E_{\beta} - E_{\alpha}$) of selected hexoses: (\square) MM3(00), (\square) B3LYP / 6-311++G(2d,2p)//B3LYP / 6-31G** 5d, and (\blacksquare) B3LYP / 6-31G** 5d.

TABLE V.
Comparative Energy Results for Water Dimer and D-Hexoses at Various Levels of Theory.

Theory	Linear Water Dimer BSSE	D-Hexoses ^a				
		Mean Values of ΔE ($\beta - \alpha$) _{QM}	RMS Errors of $\Delta E_{\text{MM3(00)}} - \Delta E_{\text{QM}}$	Mean Errors of $\Delta E_{\text{MM3(00)}} - \Delta E_{\text{QM}}$	RMS Errors of $\Delta E(\beta - \alpha)_{\text{MM3(00)}} - \Delta E(\beta - \alpha)_{\text{QM}}$	Mean Errors of $\Delta E(\beta - \alpha)_{\text{MM3(00)}} - \Delta E(\beta - \alpha)_{\text{QM}}$
B3LYP / 6-31G** 5d	-2.45	1.62	3.29	1.35	3.33	-1.99
B3LYP / 6-311++G(2d,2p) //						
B3LYP / 6-31G** 5d	-0.35	0.87	1.76	0.20	2.15	-1.23
B3LYP / AUG-cc-pVQZ ^b	0.00	0.75	1.51	0.01	1.95	-1.10

All values are in kilocalories per mole.
^aAll D-hexoses (82 conformers) used in the statistical calculation are listed in Table III.
^bStatistical results for D-hexoses at B3LYP / AUG-cc-pVQZ theory are estimated according to the trend of the BSSE in linear water dimer calculations.

bonds occur in a typical monosaccharide, the cumulative error with such systems is sufficient for the results of such calculations to be very misleading.

The agreement between QM and MM results can probably be further improved if a still higher level of theory is used. Table V shows the agreement between QM and MM results improved as the BSSE decreased in the linear water dimer calculation. There still remains a systematic discrepancy overall in the anomeric energy difference of about 1 kcal/mol between quantum mechanics and MM3. In summary, this work shows the importance of selecting an adequate basis set for the QM studies of energies of carbohydrates.

References

1. Angyal, S. J. *Aust J Chem* 1968, 21, 2737.
2. Brady, J. W. *J Am Chem Soc* 1989, 111, 5155.
3. Ha, S.; Gao, J.; Tidor, B.; Brady, J. W.; Karplus, M. *J Am Chem Soc* 1991, 113, 1553.
4. Polavarapu, P. L.; Ewig, C. S. *J Comput Chem* 1992, 13, 1255.
5. Cramer, C. J.; Truhlar, D. G. *J Am Chem Soc* 1993, 115, 5745.
6. Barrows, S. E.; Dulles, F. J.; Cramer, C. J.; French, A. D.; Truhlar, D. G. *Carbohydr Res* 1995, 276, 219.
7. Edward, J. T. *Chem Ind (London)* 1955, 1102.
8. Lemieux, R. U. In *Molecular Rearrangements*; De Mayo, P., Ed.; Interscience: New York, 1964.
9. Tvaroška, I.; Pérez, S. *Carbohydr Res* 1986, 149, 389.
10. Juaristi, E.; Cuevas, G. *Tetrahedron* 1992, 48, 5019.
11. Nørskov-Lauritsen, L.; Allinger, N. L. *J Comput Chem* 1984, 5, 326.
12. Stoddart, J. F. In *MTP International Reviews of Science. Organic Chemistry, Series One, Version 7. 1. Carbohydrates*; Aspinall, G. O., Ed.; Butterworths: London, 1973; p 1–30.
13. Tvaroška, I.; Bleha, T. In *Advances in Carbohydrate Chemistry and Biochemistry*; Tipson, R. S.; Horton, D., Eds.; Academic: San Diego, 1989; Vol. 47, p 45.
14. This value is taken from the energy difference between axial- and equatorial-2-methoxytetrahydropyran. It is about 0.75 kcal/mol with the B3LYP/6-311++G(2d,2p) theory and 1.05 kcal/mol from a temperature-dependent study of the NMR spectrum.
15. Ma, B.; Schaefer III, H. F.; Allinger, N. L. *J Am Chem Soc* 1998, 120, 3411.
16. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T. A.; Peterson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J. P.; Stewart, J. J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94, Revision E.2*; Gaussian Inc.: Pittsburgh, PA.
17. A set of five d-orbital polarization functions was used in place of the default set of six d-orbital polarization functions in the 6-31G** basis, because the results are almost the same with either basis set, and the smaller basis set is preferred for calculations on sugars because of CPU-time concerns.
18. Boys, S. F.; Bernardi, F. *Mol Phys* 1970, 19, 553.
19. (a) Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *J Am Chem Soc* 1989, 111, 8551; (b) Lii, J.-H.; Allinger, N. L. *J Am Chem Soc* 1989, 111, 8566; (c) Lii, J.-H.; Allinger, N. L. *J Am Chem Soc* 1989, 111, 8576, and subsequent articles. The MM3 program is available to all users from Tripos Associates, 1699 South Hanley Road, St. Louis, MO 63144, and to academic users only from the Quantum Chemistry Program Exchange, Indiana University, Bloomington, IN 47405.
20. Reiling, S.; Brickmann, J.; Schlenkrich, M.; Bopp, P. A. *J Comput Chem* 1996, 17, 133.
21. (a) Gobbie, H. A.; Burroughs, W. J.; Chamberlain, J.; Harries, J. E.; Jones, R. G. *Nature (Lond)* 1969, 221, 143; (b) Dyke, T. R.; Muentner, J. S. *J Chem Phys* 1974, 60, 2929; (c) Dyke, T. R.; Mack, K. M.; Muentner, J. S. *J Chem Phys* 1977, 66, 498; (d) Odutola, J. A.; Dyke, T. R. *J Chem Phys* 1980, 72, 5062.
22. (a) Matsuoka, O.; Clementi, E.; Yoshimine, M. *J Chem Phys* 1976, 64, 1351; (b) Latajka, Z.; Scheiner, S. *J Chem Phys* 1986, 84, 341; (c) Del Bene, J. E. *J Chem Phys* 1987, 86, 2110; (d) van Duijneveldt-van de Rijdt, J. G. C. M.; van Duijneveldt, F. B. *J Chem Phys* 1992, 97, 5019; (e) Saebo, S.; Tong, W.; Pulay, P. *J Chem Phys* 1993, 98, 2170; (f) Chakravorty, S. J.; Davidson, E. R. *J Phys Chem* 1993, 97, 6373; (g) Smith, B. J.; Swanton, D. J.; Pople, J. A.; Schaefer III, H. F.; Radom, L. *J Chem Phys* 1990, 92, 1240; (h) Frisch, M. J.; DelBene, J. E.; Binkley, J. S.; Schaefer III, H. F. *J Chem Phys* 1986, 84, 2279; (i) Szalewicz, K.; Cole, S. J.; Kolos, W.; Bartlett, R. J. *J Chem Phys* 1988, 89, 3662.
23. Lii, J.-H.; Allinger, N. L. *J Phys Org Chem* 1994, 7, 591.
24. Lii, J.-H. In *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R.; Allinger, N. L.; Clark, T.; Gasteiger, J.; Kollman, P. A.; Schaefer III, H. F.; Schreiner, P. R., Eds.; Wiley: Chichester, U.K., 1998; p 1271.
25. Lii, J.-H.; Chen, K.-H.; Allinger, N. L. Manuscript in preparation.
26. MM3(96) and MM3(00) force fields for carbohydrates are the same, except that one torsion parameter, O—C—O—H (type 6-1-6-21), has been updated in the later force field to fit to an *ab initio* energy profile.
27. Allinger, N. L.; Rahman, M.; Lii, J.-H. *J Am Chem Soc* 1990, 112, 8293.